

PRIORITY DOCUMENT

Jc986 US PTO  
09/9957  
11/23/01



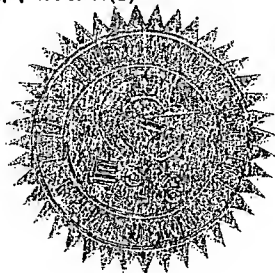
별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

출원 번호 : 특허출원 2001년 제 44269 호  
Application Number PATENT-2001-0044269

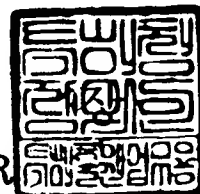
출원 년 월 일 : 2001년 07월 23일  
Date of Application JUL 23, 2001

출원인 : 한국전자통신연구원  
Applicant(s) KOREA ELECTRONICS & TELECOMMUNICATIONS RESEARCH INST



2001 년 09 월 04 일

특 허 청  
COMMISSIONER



## 【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0002
【제출일자】	2001.07.23
【발명의 명칭】	단어 추출을 이용한 특허지도 작성 시스템 및 그 방법
【발명의 영문명칭】	System for drawing patent map using technical field word, its method
【출원인】	
【명칭】	한국전자통신연구원
【출원인코드】	3-1998-007763-8
【대리인】	
【명칭】	특허법인 신성
【대리인코드】	9-2000-100004-8
【지정된변리사】	변리사 정지원, 변리사 원석희, 변리사 박해천
【포괄위임등록번호】	2000-051975-8
【발명자】	
【성명의 국문표기】	원정욱
【성명의 영문표기】	WON, Jeong Wook
【주민등록번호】	581101-1489919
【우편번호】	305-330
【주소】	대전광역시 유성구 지족동 894 열매마을 305-903
【국적】	KR
【발명자】	
【성명의 국문표기】	이형복
【성명의 영문표기】	LEE, Hyoung Bok
【주민등록번호】	600328-1468621
【우편번호】	305-333
【주소】	대전광역시 유성구 어은동 99번지 한빛아파트 126-1202
【국적】	KR

**【발명자】****【성명의 국문표기】**

고재상

**【성명의 영문표기】**

KOH, Jai Sang

**【주민등록번호】**

560806-1552718

**【우편번호】**

305-333

**【주소】**대전광역시 유성구 어은동 99번지 한빛아파트  
128-101**【국적】**

KR

**【심사청구】**

청구

**【취지】**특허법 제42조의 규정에 의하여 위와 같이 출원합  
니다. 대리인  
특허법인 신성 (인)**【수수료】****【기본출원료】**

20 면 29,000 원

**【가산출원료】**

18 면 18,000 원

**【우선권주장료】**

0 건 0 원

**【심사청구료】**

14 항 557,000 원

**【합계】**

604,000 원

**【감면사유】**

정부출연연구기관

**【감면후 수수료】**

302,000 원

**【첨부서류】**

1. 요약서·명세서(도면)\_1통

**【요약서】****【요약】****1. 청구범위에 기재된 발명이 속하는 기술분야**

본 발명은 단어 추출을 이용한 특허지도 작성 시스템 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것임.

**2. 발명이 해결하고자 하는 기술적 과제**

본 발명은, 특허 자료에서 불필요한 단어를 제거한 중요 단어의 가중치를 계산하여 특허지도를 작성하기 위한 단어를 추출하고, 이 추출된 단어를 특허와 매칭시켜 특허 지도를 작성하기 위한 단어 추출을 이용한 특허지도 작성 시스템 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공함에 그 목적이 있음.

**3. 발명의 해결방법의 요지**

본 발명은, 적어도 하나의 특허정보 제공 사이트로부터 특허 정보를 다운로드받아 저장하기 위한 저장 수단; 적어도 하나의 단어 추출 대상을 선택하고, 상기 저장 수단에 저장된 특허 정보에서 상기 선택된 단어 추출 대상의 문장을 추출하기 위한 추출 대상 선택 및 문장 추출 수단; 상기 추출 대상 선택 및 문장 추출 수단에서 선택된 추출 대상의 문장을 어절 단위로 분리하기 위한 어절 분리 수단; 상기 어절 분리 수단에서 분리된 어절 단위의 문장에서 기 저장된 불용 단어 리스트에 해당하는 불용 문자를 제거하여 불용 문자가 제거된 어절 단위의 문

장을 추출하기 위한 불용문자 제거 수단; 상기 불용문자 제거 수단에서 불용 문자가 제거된 어절 단위의 문장 또는 어절 분리 수단에서 어절 단위로 분리된 문장을 선택적으로 입력받아 단어의 갯수를 카운팅하고, 단어별로 가중치 및 가중치 합을 계산하여 단어를 추출하기 위한 단어 추출 수단; 상기 단어 추출 수단에서 추출된 단어를 특허와 매칭시키기 위한 단어 매칭 수단; 상기 단어 매칭 수단에서 매칭된 자료를 참조하여 특허지도를 작성하기 위한 특허지도 작성 수단; 및 상기 추출 대상 선택 및 문장 추출 수단의 단어 추출 대상을 제어하고, 상기 어절 분리 수단의 선택적인 출력을 제어하기 위한 특허지도 작성 제어 수단을 포함한다.

#### 4. 발명의 중요한 용도

본 발명은 특허 지도 작성 등에 이용됨.

#### 【대표도】

도 1

#### 【색인어】

특허 지도, 단어 추출, 불용 문자, 어절 분리

**【명세서】****【발명의 명칭】**

단어 추출을 이용한 특허지도 작성 시스템 및 그 방법{System for drawing patent map using technical field word, its method}

**【도면의 간단한 설명】**

도 1은 본 발명에 따른 단어 추출을 이용한 특허지도 작성 시스템의 일 실시예 구성도.

도 2는 도 1에 도시된 불용 문자 제거 모듈의 일 실시예 상세 구성도.

도 3은 도 1에 도시된 단어 추출 모듈의 일 실시예 상세 구성도.

도 4는 본 발명에 따른 단어 추출을 이용한 특허지도 작성 방법에 대한 일 실시예 흐름도.

도 5는 도 4에 도시된 불용문자 제거 과정에 대한 일 실시예 상세 흐름도.

도 6은 도 4에 도시된 단어 추출대상에서 단어 추출 과정에 대한 일 실시예 상세 흐름도.

**\* 도면의 주요부분에 대한 부호의 설명**

30 : 내부 데이터베이스      40 : 특허지도 작성 제어 모듈

50 : 추출대상 선택 및 문장 추출 모듈

60 : 어절 분리 모듈          70 : 불용문자 제거 모듈

80 : 단어 추출 모듈      90 : 단어 매칭 모듈

90 : 특허지도 작성 모듈

**【발명의 상세한 설명】**

**【발명의 목적】**

**【발명이 속하는 기술분야 및 그 분야의 종래기술】**

<13>      본 발명은 단어 추출을 이용한 특허지도 작성 시스템 및 그 방법에 관한 것으로, 보다 상세하게는 특허 자료에서 불필요한 단어를 제거한 중요 단어의 가중치를 계산하여 특허 지도를 작성하기 위한 단어 추출을 이용한 특허지도 작성 시스템 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것이다.

<14>      현재 한국에는 WIPS라는 특허정보시스템에서 특허 지도(Patent Map)를 작성하는 소프트웨어 도구(Tool)가 개발되었고, 이와 유사한 기능을 가진 것으로 한국특허청에서 PIAS(Patent Information Analysis System)를 개발하였다. 이와같이 특허정보를 소프트웨어적으로 처리하여 특허지도를 작성하는 시스템은 유료 또는 무료로 특허 정보를 제공하는 사이트의 특허 데이터베이스에 연결되어, 이 특허 데이터베이스에서 필요한 정보(서지적 사항, 청구범위, 초록 및 발명의 상세설명 등)를 자동으로 다운로드받아 이를 이미 만들어진 특허지도 작성 시스템의 데이터베이스에 저장하도록 한다.

- <15> 예를 들어, 좀더 자세히 설명하면 미국특허청(USPTO)의 특허 데이터베이스에 접속하여 원하는 검색 조건, 예를들어 검색어로 'QPSK'라는 단어를 포함한 해당 특허를 검색하면, 'QPSK'라는 단어를 포함한 특허 리스트가 웹사이트 화면에 나타난다. 이때, 특허 검색 사용자는 해당 리스트를 확인하고, 특허 리스트에 해당하는 특허 정보를 자동적으로 다운로드받아 내부 데이터베이스에 저장한다.
- <16> 그리고, 이 특허지도 작성시스템의 데이터베이스에 저장된 특허는 일련의 작업을 통하여 동일한 특허를 삭제하거나 대체시키는 작업을 통하여 필요한 특허만을 저장하게 된다.
- <17> 특허지도 작성시스템의 일종인 PIAS에서는 특허 기술별 분류 작업을 하려면 발명의 내용을 보지 않고는 분류할 수가 없었다. 따라서, 각 특허에 대하여 초록 또는 명세서를 읽어보고 특허마다 자신이 분류한 기술 분류에 할당하게 하는 방법을 취하고 있다. 예를 들어, PSK(Phase Shift Keying)라는 위상 변조 방법이 있는데, 이 PSK(Phase Shift Keying)라는 위상 변조 방법에 관한 특허들을 DPSK, QPSK 및 OPSK 등의 기술별로 분류하고자 하면, 해당 명세서들을 모두 읽어보고 각 특허에 대하여 일일이 기술 분류한 후 원하는 형태의 기술분류를 하여 여러 가지 형태의 특허 지도를 만들 수 있었다.
- <18> 따라서, 종래의 특허 작성 시스템은 특허 지도 작업을 수행하는 자가 일일이 명세서를 보아야하는 번거로움이 있고 이 때문에 특허 지도의 작성에 많은 시간이 소요되어 특허 지도의 활용도가 저하되는 문제점이 있다.



**【발명이 이루고자 하는 기술적 과제】**

<19> 이에 본 발명은, 상기와 같은 문제점을 해결하기 위해 제안된 것으로, 특허 자료에서 불필요한 단어를 제거한 중요 단어의 가중치를 계산하여 특허지도를 작성하기 위한 단어를 추출하고, 이 추출된 단어를 특허와 매칭시켜 특허 지도를 작성하기 위한 단어 추출을 이용한 특허지도 작성 시스템 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공함에 그 목적이 있다.

**【발명의 구성 및 작용】**

<20> 상기와 같은 목적을 달성하기 위한 본 발명의 장치는, 단어 추출을 이용한 특허지도 작성 시스템에 있어서, 적어도 하나의 특허정보 제공 사이트로부터 특허 정보를 다운로드받아 저장하기 위한 저장 수단; 적어도 하나의 단어 추출 대상을 선택하고, 상기 저장 수단에 저장된 특허 정보에서 상기 선택된 단어 추출 대상의 문장을 추출하기 위한 추출 대상 선택 및 문장 추출 수단; 상기 추출 대상 선택 및 문장 추출 수단에서 선택된 추출 대상의 문장을 어절 단위로 분리하기 위한 어절 분리 수단; 상기 어절 분리 수단에서 어절 단위로 분리된 문장에서 단어의 갯수를 카운팅하고, 단어별로 가중치 및 가중치 합을 계산하여 단어를 추출하기 위한 단어 추출 수단; 상기 단어 추출 수단에서 추출된 단어를 특허와 매칭시키기 위한 단어 매칭 수단; 상기 단어 매칭 수단에서 매칭된 자료를 참조하여 특허지도를 작성하기 위한 특허지도 작성 수단; 및 상기 추출 대상 선택 및

문장 추출 수단의 단어 추출 대상을 제어하기 위한 특허지도 작성 제어 수단을 포함하는 것을 특징으로 한다.

<21> 또한, 본 발명의 다른 장치는, 단어 추출을 이용한 특허지도 작성 시스템에 있어서, 적어도 하나의 특허정보 제공 사이트로부터 특허 정보를 다운로드받아 저장하기 위한 저장 수단; 적어도 하나의 단어 추출 대상을 선택하고, 상기 저장 수단에 저장된 특허 정보에서 상기 선택된 단어 추출 대상의 문장을 추출하기 위한 추출 대상 선택 및 문장 추출 수단; 상기 추출 대상 선택 및 문장 추출 수단에서 선택된 추출 대상의 문장을 어절 단위로 분리하기 위한 어절 분리 수단; 상기 어절 분리 수단에서 분리된 어절 단위의 문장에서 기 저장된 불용 단어 리스트에 해당하는 불용 문자를 제거하여 불용 문자가 제거된 어절 단위의 문장을 추출하기 위한 불용문자 제거 수단; 상기 불용문자 제거 수단에서 불용 문자가 제거된 어절 단위의 문장 또는 어절 분리 수단에서 어절 단위로 분리된 문장을 선택적으로 입력받아 단어의 갯수를 카운팅하고, 단어별로 가중치 및 가중치 합을 계산하여 단어를 추출하기 위한 단어 추출 수단; 상기 단어 추출 수단에서 추출된 단어를 특허와 매칭시키기 위한 단어 매칭 수단; 상기 단어 매칭 수단에서 매칭된 자료를 참조하여 특허지도를 작성하기 위한 특허지도 작성 수단; 및 상기 추출 대상 선택 및 문장 추출 수단의 단어 추출 대상을 제어하고, 상기 어절 분리 수단의 선택적인 출력을 제어하기 위한 특허지도 작성 제어 수단을 포함하는 것을 특징으로 한다.

<22> 한편, 본 발명의 방법은, 특허지도 작성 시스템에 적용되는 특허지도 작성 방법에 있어서, 적어도 하나의 특허정보 제공 사이트로부터 특허 정보를 다운로

드받아 내부의 데이터베이스에 저장하는 제 1 단계; 상기 내부의 데이터베이스에 저장된 특허 정보에서 적어도 하나의 단어 추출 대상을 선택하여 문장을 추출하는 제 2 단계; 상기 선택된 추출 대상의 문장을 어절 단위로 분리하는 제 3 단계; 상기 어절 단위로 분리된 문장에서 기 저장된 불용 단어 리스트에 해당하는 불용 문자를 제거하여 불용 문자가 제거된 어절 단위의 문장을 추출하는 제 4 단계; 상기 불용 문자가 제거된 어절 단위의 문장 또는 상기 어절 단위로 분리된 문장을 선택적으로 제공받아 단어의 갯수를 카운팅하고, 단어별로 가중치 및 가중치 합을 계산하여 단어를 추출하는 제 5 단계; 및 상기 추출된 단어를 특허와 매칭시켜 특허 지도를 작성하는 제 6 단계를 포함하는 것을 특징으로 한다.

<23> 한편, 본 발명은, 프로세서를 구비한 특허지도 작성 시스템에, 적어도 하나의 특허정보 제공 사이트로부터 특허 정보를 다운로드받아 내부의 데이터베이스에 저장하는 제 1 기능; 상기 내부의 데이터베이스에 저장된 특허 정보에서 적어도 하나의 단어 추출 대상을 선택하여 문장을 추출하는 제 2 기능; 상기 선택된 추출 대상의 문장을 어절 단위로 분리하는 제 3 기능; 상기 어절 단위로 분리된 문장에서 기 저장된 불용 단어 리스트에 해당하는 불용 문자를 제거하여 불용 문자가 제거된 어절 단위의 문장을 추출하는 제 4 기능; 상기 불용 문자가 제거된 어절 단위의 문장 또는 상기 어절 단위로 분리된 문장을 선택적으로 제공받아 단어의 갯수를 카운팅하고, 단어별로 가중치 및 가중치 합을 계산하여 단어를 추출하는 제 5 기능; 및 상기 추출된 단어를 특허와 매칭시켜 특허 지도를 작성하는 제 6 기능을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

- <24> 여기서 상술된 목적, 특징들 및 장점은 첨부된 도면과 관련한 다음의 상세한 설명을 통하여 보다 분명해 질 것이다. 이하 첨부된 도면을 참조하여 본 발명에 따른 바람직한 일실시예를 상세히 설명한다.
- <25> 도 1은 본 발명에 따른 단어 추출을 이용한 특허지도 작성 시스템의 일실시예 구성도이다.
- <26> 도 1에 도시된 바와 같이, 특허지도 작성 시스템은 내부 데이터베이스(30), 특허지도 작성 제어 모듈(40), 추출대상 선택 및 문장 추출 모듈(50), 어절 분리 모듈(60), 불용문자 제거 모듈(70), 단어 추출 모듈(80), 단어 매칭 모듈(90) 및 특허지도 작성 모듈(95)을 포함한다.
- <27> 상기 특허지도 작성 제어 모듈(40)은 입력되는 사용자의 선택 신호에 따라 특허지도 작성 시스템의 전반적인 특허 지도 작성을 제어하며, 추출대상 선택 및 문장 추출 모듈(50)의 추출 대상 및 단어 추출 모듈(80)의 단어 추출을 제어한다.
- <28> 또한, 상기 특허지도 작성 제어 모듈(40)은 입력되는 사용자의 선택 신호에 따라 어절 분리 모듈(60)이 단어 추출 모듈(80) 또는 불용문자 제거 모듈(70)로 어절 단위로 분리된 문장을 선택적으로 출력하도록 제어한다.
- <29> 상기 내부 데이터베이스(30)는 각 특허정보 제공 사이트(10,11,12,13)로부터 제공되는 특허 정보를 각 연동 장치(20,21,22,23)를 통해 다운로드받아 저장하고, 추출 대상 선택 및 문장 추출 모듈(50)로 특허 정보를 제공한다.

- <30> 여기서, 상기 내부 데이터베이스(30)에 저장되는 특허 정보는 서지적 사항을 포함하고, 단어를 포함하는 발명의 명칭, 초록, 청구항 및 발명의 상세 설명을 포함하며, 이를 각 추출 대상으로 구분하여 저장한다.
- <31> 상기 추출 대상 선택 및 문장 추출 모듈(50)은 내부 데이터베이스(30)로부터 제공되는 특허 정보를 토대로 해당 추출 대상을 선택하고, 추출 대상에 포함된 문장을 추출하여 어절 분리 모듈(60)로 제공한다.
- <32> 여기서, 추출 대상의 문장 중 전형적으로 사용될 수 있는 추출 대상은 청구항이 바람직하며, 발명의 상세한 설명은 해당 문장을 추출하는데 상대적으로 많은 시간이 소요되어 느린 속도의 컴퓨터로 처리하기에는 부하가 많이 걸릴 수 있다. 그리고, 보통 한가지의 추출 대상을 선택하지만 사용자의 필요에 따라 2개 이상(발명의 명칭, 청구항)의 항목을 추출 대상으로 선택할 수 있다.
- <33> 상기 어절 분리 모듈(60)은 선택된 추출 대상의 문장을 어절 단위로 분리하고, 어절 단위로 분리된 문장을 특허지도 작성 제어 모듈(40)로부터 입력되는 사용자의 선택 신호에 의해 불용문자 제거 모듈(70) 또는 단어 추출 모듈(80)로 제공한다.
- <34> 즉, 상기 어절 분리 모듈(60)은 추출 대상의 문장을 어절 단위로 분리함에 있어 공백(SPACE)으로 분리하거나, 특수문자(예를 들어 '/', 콜론, 세미콜론) 등을 포함하여 분리하는데, 여기서 어절 분리 조건은 사용자의 지정에 따라 선택되도록 한다.

- <35>      상기 불용문자 제거 모듈(70)은 어절 분리 모듈(60)로부터 제공되는 어절 단위로 분리된 문장을 기 저장된 불용문자 리스트와 비교를 통하여 문장중에서 불용 문자를 제거하고, 불용 문자가 제거된 문장을 단어 추출 모듈(80)로 제공한다.
- <36>      여기서, 상기 불용문자 제거 모듈(70)에 저장된 불용단어 리스트는 사전적인 형태로 이루어져 있어, 입력되는 어절 단위의 문장을 불용단어 리스트와 비교하여, 불용단어 리스트에 해당하는 단어가 어절 단위의 문장에 존재하면, 해당 단어는 어절 단위의 문장에서 제거된다.
- <37>      또한, 정확하게 불용 단어를 제거하기 위해서는 각 언어상의 특성을 고려하여야 한다. 따라서, 모든 언어 영역에 해당하는 공통 불용단어 리스트, 한국어에 해당하는 불용단어 리스트 및 외국어에 해당하는 불용단어 리스트를 각각 두고, 해당 어절 단위의 문장을 이 리스트들과 순차적으로 비교함으로써 최종적으로 불필요한 문자가 모두 제거된 문장을 추출함이 바람직하다. 여기서, 불용문자 제거의 세부 동작은 추후에 설명될 도 2를 토대로 더욱 자세히 설명하기로 한다.
- <38>      상기 단어 추출 모듈(80)은 특허지도 작성 제어 모듈(40)로부터 입력되는 사용자 선택에 의해 불용문자 제거 모듈(70) 또는 어절 분리 모듈(60)로부터 어절 단위의 문장을 제공받고, 이 문장에서 유용한 다수의 단어를 추출하고, 이 추출된 다수의 단어중 단어 갯수가 가장 많은 것을 기준으로 하여 나머지 다수의 단어를 나누어서 단어별 가중치를 계산하며, 계산된 단어의 가중의 합을 계산하여 단어를 추출한다.

- <39> 여기서, 상기 단어 추출 모듈(80)은 불용문자 제거 모듈(70)이 본 발명의 시스템에 적용되지 않을 경우에는 어절 분리모듈(60)로부터 출력되는 어절 단위의 문장만을 제공받으며, 한편 어절 분리 모듈(60)로부터 단어 추출 모듈(80)로 어절 단위의 문장이 제공되지 않는 경우에는 불용문자 제거 모듈(70)로부터 불용문자가 제거된 어절 단위의 문장만을 제공받는다. 이 두 경우에는 상기 특허지도 작성 제어 모듈(40)의 선택 신호가 필요없다.
- <40> 상기 단어 매칭 모듈(90)은 단어 추출 모듈(80)에서 추출된 단어를 특허와 매칭시키고, 상기 매칭된 단어 정보를 특허지도 작성 모듈(95)로 제공한다.
- <41> 여기서, 상기 단어 매칭과정은 추출된 단어와 특허를 1 : 1로 매칭시키는게 일반적이나, 사용자에 의해 1 : N의 복수 매칭도 가능하다. 예를 들면, 한 특허가 단어A 및 단어B를 2 개 이상 포함하는 경우에 단어A가 많은지 단어B가 많은지를 비교해서 1 : 1 매칭으로 처리할 수 있고, 동시에 두 특허이상을 포함하는 1 : N의 복수 매칭을 할 수 있도록 한다.
- <42> 상기 특허지도 작성 모듈(95)은 단어 매칭 모듈(90)로부터 매칭된 특허를 토대로 기술분야별, 국가별, 출원기간별 및 출원인별로 특허 지도를 작성한다.
- <43> 예를들어, '위상 변조'에 해당하는 특허를 모아다가 PSK 변조방식에서 'BPSK', 'QPSK' 및 'OCQPSK' 등으로 구분하여 분류할 수 있으며, 다시 'QPSK'를 다시 분류하고자 하면 '변조' 및 '복조'로 분류할 수 있는데, 이때에는 현재 내부 데이터베이스(30)에 매칭되어 입력된 데이터를 가지고 재 분류 과정을 통해 가능하며, 반복적인 작업을 통해 특허 지도 작성에 따른 기술 분류를 더욱 세분화 할 수 있다.

- <44> 도 2는 도 1에 도시된 불용문자 제거 모듈의 일실시에 상세 구성도이다.
- <45> 도 2에 도시된 바와 같이, 상기 불용문자 제거 모듈(70)은 공통 불용문자 제거부(72), 한국어 불용문자 제거부(75), 외국어 불용문자 제거부(78)를 포함한다. 그리고, 상기 한국어 불용문자 제거부(75)는 불용 첫음절 제거부(73) 및 불용 끝음절 제거부(74)를 포함하고, 상기 외국어 불용문자 제거부(78)는 복수형 제거부(76) 및 대소문자 동일 처리부(77)를 포함한다.
- <46> 상기 공통 불용문자 제거부(72)는 세계 모든 언어에서 적용되는 공통 불용 단어 리스트를 토대로 어절 분리 모듈(60)에서 제공되는 어절 단위의 문장에서 불필요한 단어를 제거하고, 공용 불용 단어가 제거된 문장을 한국어 불용문자 제거부(75)의 불용 첫음절 제거부(73)로 제공한다.
- <47> 즉, 공용 불용문자 제거부(72)는 어절 단위의 문장을 기 저장된 공통 불용 단어 리스트와 비교하여, 어절 단위의 문장중에 공통 불용단어 리스트에 포함된 단어를 제거하는 것이다.
- <48> 예를들어, 상기 공용 불용단어 리스트에는 한국어의 경우에는 '잘', '매우', '그리고' 및 '또는' 등과 같은 불용 단어가, 영어의 경우에는 'a', 'the', 'this', 'that', 'much' 및 'little' 등과 같은 단어가, 독일어의 경우에는 'des', 'dem', 'den' 등과 같은 단어가, 그리고 프랑스어에는 'La' 및 'De' 등과 같은 단어가 저장된다.
- <49> 따라서, 공통 불용문자 제거부(72)는 상기 어절 분리 모듈(60)로부터 제공되는 어절 단위의 문장중에 공용 불용단어 리스트에 해당하는 단어를 제거하고,



이 단어가 제거된 문장을 한국어 불용문자 제거부(75)의 불용 첫음절 제거부(73)로 제공한다.

<50>      상기 불용 첫음절 제거부(73)는 공통 불용문자 제거부(72)로부터 제공되는 공통 불용문자가 제거된 어절 단위의 문장에서 첫 음절의 불용 단어를 제거하고, 이 첫 음절의 불용 단어가 제거된 문장을 불용 끝음절 제거부(74)로 제공한다.

<51>      즉, 상기 불용 첫음절 제거부(73)는 공통 불용 문자가 제거된 어절 단위의 문장중에 첫음절 불용단어 리스트에 해당하는 단어를 제거하는 것으로, 예를들어, 상기 첫음절 불용단어 리스트에는 '극', '최'와 같은 접두어를 포함하고 있어, 입력되는 어절 단위의 문장중에 첫음절 불용단어 리스트에 해당하는 단어를 제거하는 것이다.

<52>      상기 불용 끝음절 제거부(74)는 불용 첫음절 제거부(73)로부터 제공되는 첫음절의 불용 단어가 제거된 어절 단위의 문장중 끝음절 불용단어 리스트에 해당하는 불용 단어를 제거하고, 이 끝음절 불용 단어가 제거된 어절 단위의 문장을 외국어 불용문자 제거부(78)의 복수형 제거부(76)로 제공한다.

<53>      여기서, 상기 끝음절 불용단어 리스트는 목적격 조사인 '을' 및 '를', 관형격 조사인 '의', 주격 조사인 '는' 및 '은'을 포함하고 있으며, 또한, 동사의 어미에 해당하는 '하는', '하고' 및 '하여' 등의 불용 단어를 포함하고 있다. 따라서, 첫음절의 불용 단어가 제거된 어절 단위의 문장중에 끝음절 불용단어 리스트에 해당하는 단어가 제거된다.

- <54>      상기 복수형 제거부(76)는 불용 끝음절 제거부(74)에서 끝 음절의 불용 단어가 제거된 어절 단위의 문장중에 복수형 불용단어 리스트에 해당하는 불용 단어를 제거하고, 이 복수형 불용 단어가 제거된 어절 단위의 문장을 대소문자 동일 처리부(77)로 제공한다.
- <55>      여기서, 상기 복수형 불용단어 리스트는 영어에서 복수형인 's', 'es'을 포함하고, 동일한 개념으로 불어, 독어 및 일본어의 복수 형태의 불용 단어를 포함하고 있다. 따라서, 끝 음절의 불용 단어가 제거된 어절 단위의 문장중에 복수형 불용단어 리스트에 해당하는 단어가 제거된다.
- <56>      상기 대소문자 동일 처리부(77)는 복수형 제거부(76)로부터 복수형 불용 단어가 제거된 어절 단위의 문장중에 동일단어 리스트에 해당하는 단어를 동일한 단어로 처리하여, 동일 단어로 처리된 어절 단위의 문장을 단어 추출 모듈(80)로 제공한다.
- <57>      도 3은 도 1에 도시된 단어 추출 모듈의 일실시에 상세 구성도이다.
- <58>      도 3에 도시된 바와 같이, 상기 단어 추출 모듈(80)은 단어별 카운팅부(82), 가중치 계산부(83), 단어가중치 합 계산부(84), 불용단어 제거부(85) 및 단어 추출부(86)를 포함한다.
- <59>      상기 단어별 카운팅부(82)는 불용문자 제거 모듈(70)로부터 제공되는 불용문자가 제거된 어절 단위의 문장 또는 어절 분리 모듈(60)로부터 어절 단위로 분리된 문장을 입력받고, 해당 문장에서 단어별로 그 수를 카운터하고, 해당 단어별의 카운트 정보를 가중치 계산부(83)로 제공한다.

- <60>      상기 가중치 계산부(83)는 단어별 카운팅부(82)로부터 제공된 해당 단어별의 카운트 정보를 참조하여 카운터 횟수가 가장 많은 단어 즉, 문장중에 빈도수가 가장 높은 단어의 카운터 값으로 나머지의 단어들의 가중치를 계산하고, 계산된 단어별 가중치 정보를 단어가중치 합 계산부(84)로 제공한다.
- <61>      예를들어, 첫번째 특허에서는 단어A가 20번, 단어B가 15번, 단어C가 10번, 단어D가 5번 나왔다고 가정하면, 이때, 최고 빈도수 단어는 단어A이다. 이 단어A는 20번 나왔으므로 단어 빈도수가 최고 높은 값 20으로 나머지 단어들을 나누면 단어A가 1, 단어B가 0.75, 단어C가 0.5, 단어D가 0.25가 된다. 이때, 단어A의 1, 단어B의 0.75, 단어C의 0.5 및 단어D의 0.25가 가중치가 된다.
- <62>      그리고, 두 번째 특허에서 단어 빈도수를 조사하여 단어A가 5번, 단어B가 10번, 단어C가 0번, 단어D가 5번, 단어F가 5번 나왔다고 가정하면, 이때, 다시 단어 최고 빈도수 10으로 나누어 가중치를 구하면 단어A가 0.5, 단어B가 1, 단어C가 0, 단어D가 0.5, 단어F가 0.5가 된다. 따라서, 상기와 같은 방법을 각 특허들에 적용하면 해당 특허의 각 단어별 가중치 정보가 생성되고, 이 생성된 단어별 가중치 정보가 단어가중치 합 계산부(84)로 제공된다.
- <63>      상기 단어가중치 합 계산부(84)는 가중치 계산부(83)로부터 제공되는 해당 특허의 각 단어별 가중치 정보를 이용하여 단어별 가중치 정보의 합을 계산하고, 계산된 가중치의 합을 외부의 디스플레이로 표시하며, 계산된 단어별 가중치 합 정보를 특허지도 작성 제어 모듈(40)로부터 입력되는 선택 신호에 따라 불용단어 제거부(85) 또는 단어 추출부(86)로 제공한다.

- <64> 즉, 사용자가 디스플레이를 통해 표시되는 단어별 가중치 정보의 합을 확인하고, 특허지도 작성 제어모듈(40)을 통해 단순범위 선택 모드를 선택하면 가중치 순서대로 단어A, 단어B, 단어D, 단어F, 단어C의 정보에 사용자가 설정한 범위의 단어만을 단어 추출부(86)로 제공하는데, 예를들어, 사용자가 단어A, 단어B, 단어D 만을 선택하면 해당 단어만이 단어 추출부(86)로 제공된다.
- <65> 또한, 사용자가 디스플레이를 통해 표시되는 단어별 가중치 정보의 합을 확인하고, 특허지도 작성 제어 모듈(40)을 통해 불용단어 제거 모드를 선택하여 제거할 불용 단어를 선택한다. 그러면, 상기 단어가중치 합 계산부(84)는 계산된 단어별 가중치 합 정보 및 사용자가 선택한 단어 제거 정보를 불용 단어 제거부(85)로 제공한다. 이때, 불용 단어 제거부(85)는 단어별 가중치 합 정보의 단어에서 사용자가 선택한 단어를 제거하고, 나머지의 단어만을 단어 추출부(86)로 제공한다.
- <66> 예를들어, 50개의 특허를 분석한다고 할때 가중치의 합이 단어A가 40, 단어B가 20, 단어C가 7, 단어D가 20, 단어F가 10일 경우에, 가중치 순서대로 단어 A, 단어 B, 단어 D, 단어 F, 단어 C가 불용 단어 제거부(85)로 입력되면, 상기 불용 단어 제거부(85)는 사용자의 선택에 의해 가중치가 높은 단어라도 불필요한 단어라고 지정된 단어를 제거한다. 따라서, 상기 불용 단어 제거부(86)는 단어A, 단어B, 단어F가 사용자에게 의해 불용 단어로 지정되면, 나머지인 단어 B, 단어 F, 단어 C만을 단어 추출부(86)로 제공한다.

- <67>        여기서, 사용자가 단순범위 선택 모드 또는 불용단어 제거 모드를 선택적으로 설정하도록 하고 있으나, 두 모드(단순범위 선택 모드 또는 불용단어 제거 모드)를 모두 만족하는 단어만을 추출함이 더욱 바람직할 것이다.
- <68>        상기 단어 추출부(86)는 단어가중치 합 계산부(84)를 통해 입력되는 단어 가중치 합 정보 또는 불용 단어 제거부(85)를 통해 입력되는 불용 단어가 제거된 단어 가중치 합 정보를 토대로 해당 단어를 추출하여 단어 매칭 모듈(90)로 제공한다.
- <69>        도 4는 본 발명에 따른 단어 추출을 이용한 특허지도 작성 방법에 대한 일 실시예 흐름도이다.
- <70>        먼저, 각 특허정보 제공 사이트(10,11,12,13)로부터 제공되는 특허 정보를 각 연동 장치(20,21,22,23)를 통해 다운로드받아 내부 데이터베이스(30)에 저장한다. 이후, 특허지도 작성 제어모듈(40)의 제어 신호에 의해 추출 대상 선택 및 문장 추출 모듈(50)은 내부 데이터베이스(30)로부터 제공되는 특허 정보에서 해당 추출 대상을 선택하고, 추출 대상에 포함된 문장을 추출하여 어절 분리 모듈(60)로 제공한다(100).
- <71>        여기서, 상기 내부 데이터베이스(30)에 저장되는 특허 정보는 서지적 사항을 포함하고, 단어를 포함하는 발명의 명칭, 초록, 청구항 및 발명의 상세 설명을 포함하며, 이를 각 추출 대상으로 구분하여 저장한다.

- <72> 그러면, 상기 어절 분리 모듈(60)을 통해 선택된 추출 대상의 문장은 어절 단위로 분리되고(200), 어절 단위로 분리된 문장에서 불용 문자를 제거할 것인지를 판단한다(300).
- <73> 상기 과정(300)에서 판단 결과, 어절 단위로 분리된 문장에서 불용 문자를 제거할 경우에는 어절 단위로 분리된 문장을 불용문자 제거 모듈(70)로 제공하여, 불용문자 제거 모듈(70)의 불용단어 리스트를 참조하여 해당하는 불용 문자를 제거하고, 분리된 문장에서의 단어를 가중치에 따라 추출하는 과정(500)으로 진행한다.
- <74> 상기 과정(300)에서 판단 결과, 단위로 분리된 문장에서 불용 문자를 제거하지 않을 경우 분리된 문장에서의 단어를 가중치에 따라 추출하는 과정(500)으로 진행한다.
- <75> 즉, 상기 불용 문자를 제거할 경우에는 어절 분리 모듈(60)에서 어절 단위로 분리된 문장중에 불용 문자를 불용문자 제거 모듈(70)에서 제거하고, 이 불용 문자가 제거된 어절 단위의 문장을 단어 추출 모듈(80)로 제공한다. 여기서, 상기 불용문자 제거 모듈(70)은 불용단어 리스트 또는 단어 사전을 저장하고 있어, 입력되는 어절 단위의 문장을 불용단어 리스트와 비교를 통해 불필요한 문자를 제거하는데, 각 언어상의 특성이 다르기 때문에 이를 모든 언어 영역에 해당하는 불용단어 리스트, 한국어에 해당하는 불용단어 리스트 및 외국어에 해당하는 불용단어 리스트를 각각 두고, 해당 어절 단위의 문장과 이 리스트들을 순차적으로 비교함으로써 최종적으로 불필요한 문자를 제거함이 바람직하다.

<76> 한편, 상기 과정(500)에서는 불용문자 제거 모듈(70)로부터 불용 문자가 제거된 어절 단위의 문장 또는 상기 어절 분리 모듈(60)로부터의 어절 단위로 분리된 문장을 제공받고, 이 문장에서 유용한 다수의 단어를 추출하고, 이 추출된 다수의 단어중 단어 갯수가 가장 많은 것을 기준으로 하여 나머지 다수의 단어를 나누어서 단어별 가중치를 계산하고, 계산된 단어의 가중치의 합을 계산하여 단어를 추출하고(500), 상기 추출된 단어는 단어 매칭 모듈(80)에 의해 특허와 매칭되고, 상기 매칭된 정보를 토대로 특허 지도를 작성한다(600).

<77> 여기서, 추출된 단어와 특허를 1 : 1로 매칭시키는게 일반적이나, 사용자에게 의해 1 : N의 복수 매칭도 가능하다. 예를 들면, 한 특허가 단어A 및 단어B를 2개이상 포함하는 경우에 단어A가 많은지 단어B가 많은지를 비교해서 1 : 1 매칭 처리할 수 있고, 동시에 2 곳을 포함하는 1 : N의 복수 매칭도 가능하다.

<78> 또한, 특허 지도의 작성은 분류된 단어를 토대로 기술분야별, 국가별, 출원기간별 및 출원인별로 특허 지도를 작성한다.

<79> 도 5는 도 4에 도시된 불용문자 제거 과정에 대한 일실시에 상세 흐름도이다.

<80> 도 5에 도시된 바와 같이, 어절 분리 모듈(60)을 통해 어절 단위로 분리된 문장중 공통 불용문자 제거부(72)의 공통 불용단어 리스트에 해당하는 불 필요한 단어가 제거되고, 공용 불용 단어가 제거된 문장은 불용 첫음절 제거부(73)로 제공된다(410).

- <81> 그러면, 공통 불용문자가 제거된 어절 단위의 문장은 상기 불용 첫음절 제거부(73)를 통해 어절 단위의 문장에서 첫음절 불용단어 리스트에 해당하는 단어가 제거되고, 이 첫 음절의 불용 단어가 제거된 문장은 불용 끝음절 제거부(74)로 제공된다(420).
- <82> 이때, 상기 과정(420)에서 첫음절 불용단어 리스트에 해당하는 단어가 제거된 문장은 불용 끝음절 제거부(74)를 통해 끝음절 불용단어 리스트에 해당하는 단어가 제거되고, 이 끝음절 불용 단어가 제거된 어절 단위의 문장은 외국어 불용문자 제거부(78)의 복수형 제거부(76)로 제공된다(430).
- <83> 그리고, 상기 과정(430)에서 끝 음절의 불용 단어가 제거된 어절 단위의 문장중 복수형 제거부(76)를 통해 복수형 불용단어 리스트에 해당하는 불용 단어가 제거되고, 이 복수형 불용 단어가 제거된 어절 단위의 문장중 대소 문자 동일 처리부(77)를 통해 동일단어 리스트에 해당하는 단어가 동일한 단어로 처리되고(440), 동일 단어로 처리된 어절 단위의 문장을 단어 추출 모듈(80)로 제공한다.
- <84> 도 6은 도 4에 도시된 단어 추출대상에서 단어 추출 과정에 대한 일실시에 상세 흐름도이다.
- <85> 도 6에 도시된 바와 같이, 불용문자 제거 모듈(70)로부터 제공되는 불용 문자가 제거된 어절 단위의 문장 또는 어절 분리 모듈(60)로부터 어절 단위로 분리된 문장을 입력받고, 단어별 카운팅부(82)를 통해 해당 문장을 단어별로 그 수를 카운터하고, 해당 단어별의 카운트 정보를 가중치 계산부(83)로 제공한다(510).



- <86> 이후, 상기 과정(510)에서 제공되는 해당 단어별의 카운트 정보를 참조하여 카운터 횟수가 가장 많은 즉, 문장중에 빈도수가 가장 높은 단어의 카운터 값으로 나머지의 단어들의 가중치를 계산한다(520).
- <87> 이후, 상기 과정(520)에서 계산된 단어별 가중치 정보를 단어가중치 합 계산부(84)로 제공하여 단어별 가중치 정보의 합으로 계산하고, 단어별 가중치 정보의 합을 토대로 불용단어 제거 모드를 수행할 것인지, 단순범위 선택 모드를 수행할 것인지를 판단한다(540).
- <88> 상기 과정(540)에서 판단한 결과, 사용자가 디스플레이를 통해 표시되는 단어별 가중치 정보의 합을 확인하고 단순범위 선택 모드를 선택할 경우, 가중치 순서대로 각 단어의 가중치 합 정보를 단어 추출부(86)로 제공하여 사용자가 선택한 범위내의 가중치 합 정보를 추출한다(560).
- <89> 상기 과정(540)에서 판단한 결과, 단어별 가중치 정보의 합을 확인하고 불용단어 제거 모드를 선택할 경우에는, 계산된 단어별 가중치의 합 정보를 불용단어 제거부(85)로 제공한다. 따라서, 상기 불용 단어 제거부(85)는 단어별 가중치 합 정보의 단어에서 사용자가 선택한 단어를 제거후에 가중치 합 정보를 추출하고(550), 추출된 단어를 토대로 특허 지도를 작성하는 단계(600)으로 진행된다.
- <90> 상술한 바와 같은 본 발명의 방법은 프로그램으로 구현되어 컴퓨터로 읽을 수 있는 기록매체(씨디롬, 램, 롬, 플로피 디스크, 하드 디스크, 광자기 디스크 등)에 저장될 수 있다.

<91>       이상에서 설명한 본 발명은 진술한 실시예 및 첨부된 도면에 의해 한정되는 것이 아니고, 본 발명의 기술적 사상을 벗어나지 않는 범위내에서 여러가지 치환, 변형 및 변경이 가능하다는 것이 본 발명이 속하는 기술분야에서 통상의 지식을 가진자에게 있어 명백할 것이다.

**【발명의 효과】**

<92>       상기와 같은 본 발명은, 특허 자료에서 불필요한 단어를 제거한 중요 단어의 가중치를 계산하여 특허지도를 작성함으로써, 사용자의 요구에 부합되는 정확한 특허지도의 작성이 가능하다.

<93>       또한, 본 발명은, 특허지도 작업을 수행하는 자가 일일이 명세서를 보지 않아도 되므로, 특허지도 작성 시간을 획기적으로 단축시킬 수 있고, 사용자의 편리성을 크게 향상시킬 수 있다.

**【특허청구범위】****【청구항 1】**

단어 추출을 이용한 특허지도 작성 시스템에 있어서,

적어도 하나의 특허정보 제공 사이트로부터 특허 정보를 다운로드받아 저장하기 위한 저장 수단;

적어도 하나의 단어 추출 대상을 선택하고, 상기 저장 수단에 저장된 특허 정보에서 상기 선택된 단어 추출 대상의 문장을 추출하기 위한 추출 대상 선택 및 문장 추출 수단;

상기 추출 대상 선택 및 문장 추출 수단에서 선택된 추출 대상의 문장을 어절 단위로 분리하기 위한 어절 분리 수단;

상기 어절 분리 수단에서 어절 단위로 분리된 문장에서 단어의 갯수를 카운팅하고, 단어별로 가중치 및 가중치 합을 계산하여 단어를 추출하기 위한 단어 추출 수단;

상기 단어 추출 수단에서 추출된 단어를 특허와 매칭시키기 위한 단어 매칭 수단;

상기 단어 매칭 수단에서 매칭된 자료를 참조하여 특허지도를 작성하기 위한 특허지도 작성 수단; 및

상기 추출 대상 선택 및 문장 추출 수단의 단어 추출 대상을 제어하기 위한 특허지도 작성 제어 수단

을 포함하는 단어 추출을 이용한 특허지도 작성 시스템.

**【청구항 2】**

제 1 항에 있어서,

상기 단어 추출 수단은,

상기 어절 분리 수단에서 출력된 어절 단위 문장에서 단어별로 단어의 갯수를 카운팅하기 위한 단어별 카운팅부;

상기 단어별 카운팅부에서 카운팅된 단어 중 갯수가 가장 많은 단어의 수로 나머지 각 단어들의 수를 나누어 단어별로 가중치를 계산하기 위한 가중치 계산부;

상기 가중치 계산부에서 계산된 단어별 가중치의 합을 계산하기 위한 단어 가중치 합 계산부; 및

상기 단어가중치 합 계산부로부터의 단어를 제공받아 가중치 합순으로 정리하여 가중치 순위순으로 선택된 범위까지 단어를 추출하기 위한 단어 추출부를 포함하는 단어 추출을 이용한 특허지도 작성 시스템.

**【청구항 3】**

제 1 항에 있어서,

상기 단어 추출 수단은,

상기 어절 분리 수단에서 출력된 어절 단위 문장에서 단어별로 단어의 갯수를 카운팅하기 위한 단어별 카운팅부;

상기 단어별 카운팅부에서 카운팅된 단어 중 갯수가 가장 많은 단어의 수로 나머지 각 단어들의 수를 나누어 단어별로 가중치를 계산하기 위한 가중치 계산부;

상기 가중치 계산부에서 계산된 단어별 가중치의 합을 계산하기 위한 단어 가중치 합 계산부;

상기 단어가중치 합 계산부에서 계산된 가중치 합순으로 단어를 정리하여 해당 단어에서 불용 단어를 제거하기 위한 불용단어 제거부; 및

상기 불용단어 제거부에서 불용 단어가 제거된 단어 또는 상기 단어가중치 합 계산부로부터의 단어를 상기 특허지도 작성 제어 수단의 제어에 따라 선택적으로 제공받아 가중치 합순으로 정리하여 가중치 순위순으로 선택된 범위까지 단어를 추출하기 위한 단어 추출부

를 포함하는 단어 추출을 이용한 특허지도 작성 시스템.

#### 【청구항 4】

제 1 항 내지 제 3 항 중 어느 한 항에 있어서,

상기 단어 추출 대상은,

발명의 명칭, 청구범위, 초록 또는 발명의 상세한 설명 중에서 어느 한가지 이상의 대상을 포함하는 것을 특징으로 하는 단어 추출을 이용한 특허지도 작성 시스템.

## 【청구항 5】

단어 추출을 이용한 특허지도 작성 시스템에 있어서,

적어도 하나의 특허정보 제공 사이트로부터 특허 정보를 다운로드받아 저장하기 위한 저장 수단;

적어도 하나의 단어 추출 대상을 선택하고, 상기 저장 수단에 저장된 특허 정보에서 상기 선택된 단어 추출 대상의 문장을 추출하기 위한 추출 대상 선택 및 문장 추출 수단;

상기 추출 대상 선택 및 문장 추출 수단에서 선택된 추출 대상의 문장을 어절 단위로 분리하기 위한 어절 분리 수단;

상기 어절 분리 수단에서 분리된 어절 단위의 문장에서 기 저장된 불용 단어 리스트에 해당하는 불용 문자를 제거하여 불용 문자가 제거된 어절 단위의 문장을 추출하기 위한 불용문자 제거 수단;

상기 불용문자 제거 수단에서 불용 문자가 제거된 어절 단위의 문장 또는 어절 분리 수단에서 어절 단위로 분리된 문장을 선택적으로 입력받아 단어의 갯수를 카운팅하고, 단어별로 가중치 및 가중치 합을 계산하여 단어를 추출하기 위한 단어 추출 수단;

상기 단어 추출 수단에서 추출된 단어를 특허와 매칭시키기 위한 단어 매칭 수단;

상기 단어 매칭 수단에서 매칭된 자료를 참조하여 특허지도를 작성하기 위한 특허지도 작성 수단; 및

상기 추출 대상 선택 및 문장 추출 수단의 단어 추출 대상을 제어하고, 상기 어절 분리 수단의 선택적인 출력을 제어하기 위한 특허지도 작성 제어 수단을 포함하는 단어 추출을 이용한 특허지도 작성 시스템.

**【청구항 6】**

제 5 항에 있어서,

상기 불용문자 제거 수단은,

각국의 언어별 불용 단어에 적용되는 공통 불용 단어 리스트에 따라, 상기 어절 분리 수단으로부터 입력되는 어절 단위의 문장중에서 불용 문자를 제거하기 위한 공통 불용문자 제거부 ;

어절 단위의 문장중에서 첫 음절 불용단어 리스트에 따라 불용의 첫 음절을 제거하거나, 끝음절 불용단어 리스트에 따라 불용의 끝 음절을 제거하기 위한 불용문자제거부; 및

외국어의 복수형 단어를 제거하거나 대문자와 소문자를 동일하게 처리하기 위한 외국어 불용문자 제거부

중 어느 하나 이상을 포함하는 단어 추출을 이용한 특허지도 작성 시스템.

**【청구항 7】**

제 5 항 또는 제 6 항에 있어서,

상기 단어 추출 대상은,

발명의 명칭, 청구범위, 초록 또는 발명의 상세한 설명 중에서 어느 한가지 이상의 대상을 포함하는 것을 특징으로 하는 단어 추출을 이용한 특허지도 작성 시스템.

【청구항 8】

특허지도 작성 시스템에 적용되는 특허지도 작성 방법에 있어서,

적어도 하나의 특허정보 제공 사이트로부터 특허 정보를 다운로드받아 내부의 데이터베이스에 저장하는 제 1 단계;

상기 내부의 데이터베이스에 저장된 특허 정보에서 적어도 하나의 단어 추출 대상을 선택하여 문장을 추출하는 제 2 단계;

상기 선택된 추출 대상의 문장을 어절 단위로 분리하는 제 3 단계;

상기 어절 단위로 분리된 문장에서 기 저장된 불용 단어 리스트에 해당하는 불용 문자를 제거하여 불용 문자가 제거된 어절 단위의 문장을 추출하는 제 4 단계;

상기 불용 문자가 제거된 어절 단위의 문장 또는 상기 어절 단위로 분리된 문장을 선택적으로 제공받아 단어의 갯수를 카운팅하고, 단어별로 가중치 및 가중치 합을 계산하여 단어를 추출하는 제 5 단계; 및

상기 추출된 단어를 특허와 매칭시켜 특허 지도를 작성하는 제 6 단계

를 포함하는 단어 추출을 이용한 특허지도 작성 방법.



**【청구항 9】**

제 8 항에 있어서,

상기 제 4 단계는,

각국의 언어별 불용 단어에 적용되는 공통 불용 단어 리스트에 따라, 상기 어절 단위로 분리된 문장중에서 불용 문자를 제거하는 제 7 단계를 더 포함하는 단어 추출을 이용한 특허지도 작성 방법.

**【청구항 10】**

제 8 항에 있어서,

상기 제 4 단계는,

상기 공통의 불용 문자가 제거된 어절 단위의 문장에서 한국어의 첫 음절 불용단어 리스트에 따라 불용의 첫 음절을 제거하고, 끝음절 불용단어 리스트에 따라 불용의 끝 음절을 제거하는 제 8 단계

를 더 포함하는 단어 추출을 이용한 특허지도 작성 방법.

**【청구항 11】**

제 8 항에 있어서,

상기 제 4 단계는,

상기 한국어의 첫 음절 및 끝 음절의 불용 문자가 제거된 어절 단위 문장에서 외국어의 복수형 단어를 제거하고 대문자와 소문자를 동일하게 처리하는 제 9 단계

를 더 포함하는 단어 추출을 이용한 특허지도 작성 방법.

【청구항 12】

제 8 항에 있어서,

상기 제 5 단계는,

상기 불용 문자가 제거된 어절 단위의 문장 또는 상기 어절 단위로 분리된 문장을 선택적으로 제공받아 단어별로 단어의 갯수를 카운팅하는 제 7 단계;

상기 카운팅된 단어 중 갯수가 가장 많은 단어의 수로 나머지 각 단어들의 수를 나누어 단어별로 가중치를 계산하고, 이 계산된 단어별 가중치의 합을 계산하는 제 8 단계; 및

상기 가중치 합이 계산된 단어를 가중치 합순으로 정리하여 가중치 순위 순으로 선택된 범위까지 단어를 추출하는 제 9 단계

를 포함하는 단어 추출을 이용한 특허지도 작성 방법.

【청구항 13】

제 8 항에 있어서,

상기 제 5 단계는,

상기 불용 문자가 제거된 어절 단위의 문장 또는 상기 어절 단위로 분리된 문장을 선택적으로 제공받아 단어별로 단어의 갯수를 카운팅하는 제 7 단계;

상기 카운팅된 단어 중 갯수가 가장 많은 단어의 수로 나머지 각 단어들의 수를 나누어 단어별로 가중치를 계산하고, 이 계산된 단어별 가중치의 합을 계산하는 제 8 단계;

상기 가중치 합이 계산된 단어를 가중치 합순으로 정리하여 해당 단어에서 불용 단어를 제거하는 제 9 단계; 및

상기 불용 단어가 제거된 단어 또는 상기 가중치 합이 계산된 단어를 선택적으로 제공받아 가중치 합순으로 정리하여 가중치 순위 순으로 선택된 범위까지 단어를 추출하는 제 10 단계

를 포함하는 단어 추출을 이용한 특허지도 작성 방법.

#### 【청구항 14】

프로세서를 구비한 특허지도 작성 시스템에,

적어도 하나의 특허정보 제공 사이트로부터 특허 정보를 다운로드받아 내부의 데이터베이스에 저장하는 제 1 기능;

상기 내부의 데이터베이스에 저장된 특허 정보에서 적어도 하나의 단어 추출 대상을 선택하여 문장을 추출하는 제 2 기능;

상기 선택된 추출 대상의 문장을 어절 단위로 분리하는 제 3 기능;

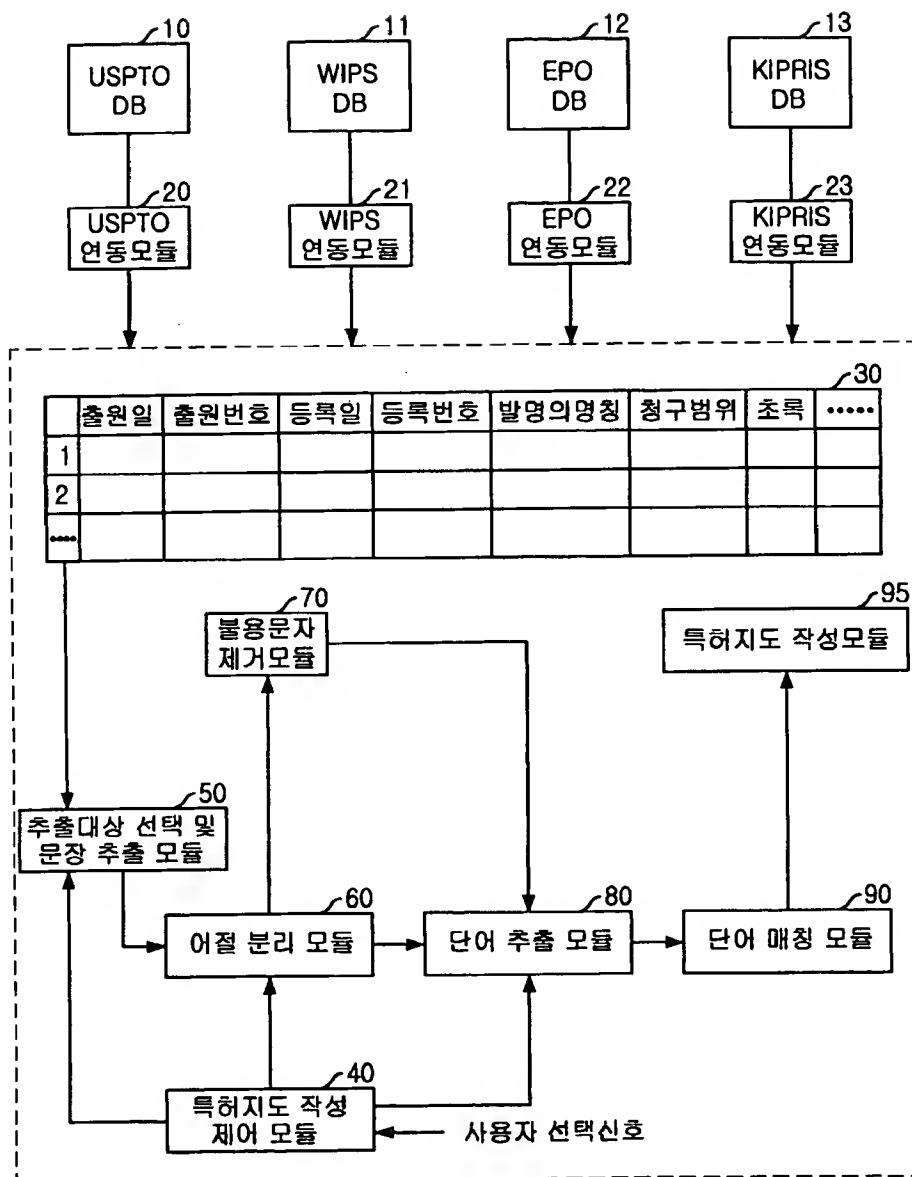
상기 어절 단위로 분리된 문장에서 기 저장된 불용 단어 리스트에 해당하는 불용 문자를 제거하여 불용 문자가 제거된 어절 단위의 문장을 추출하는 제 4 기능;

상기 불용 문자가 제거된 어절 단위의 문장 또는 상기 어절 단위로 분리된 문장을 선택적으로 제공받아 단어의 갯수를 카운팅하고, 단어별로 가중치 및 가중치 합을 계산하여 단어를 추출하는 제 5 기능; 및

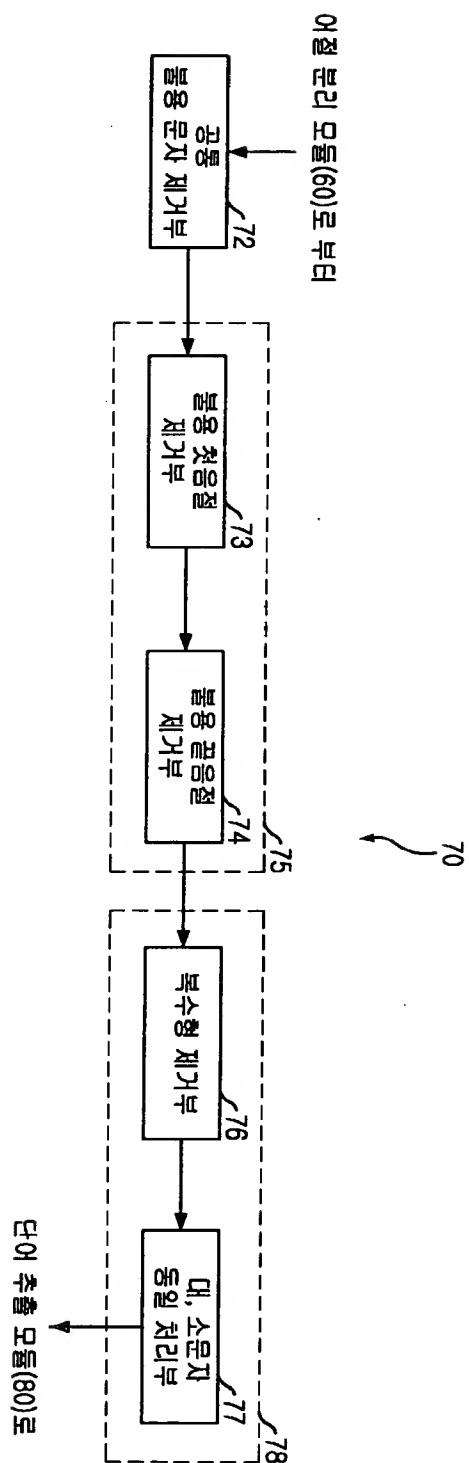
상기 추출된 단어를 특허와 매칭시켜 특허 지도를 작성하는 제 6 기능을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

【도면】

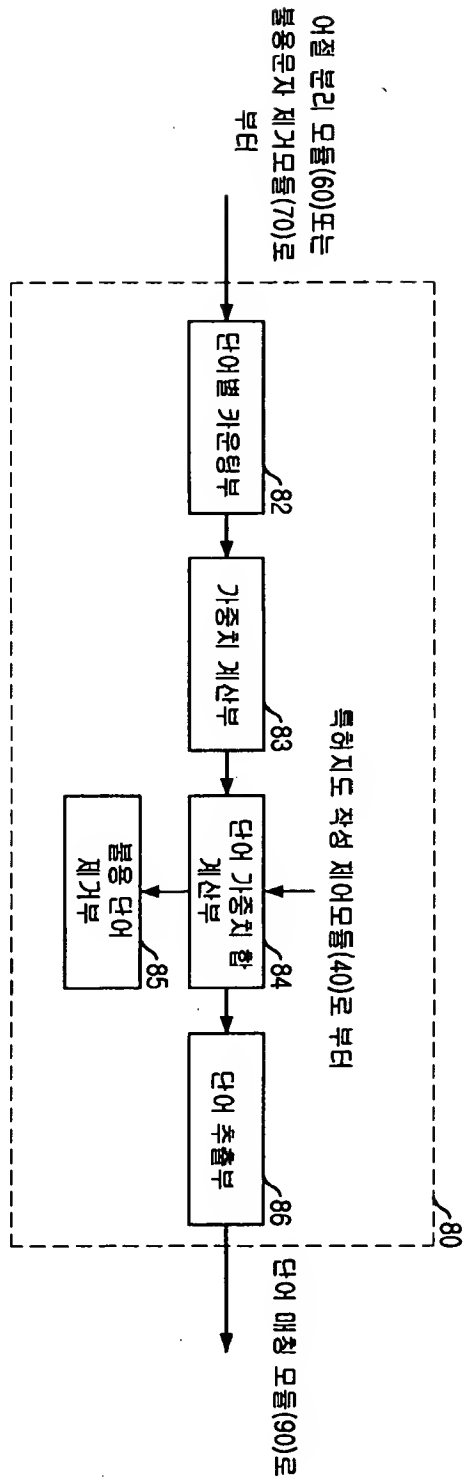
【도 1】



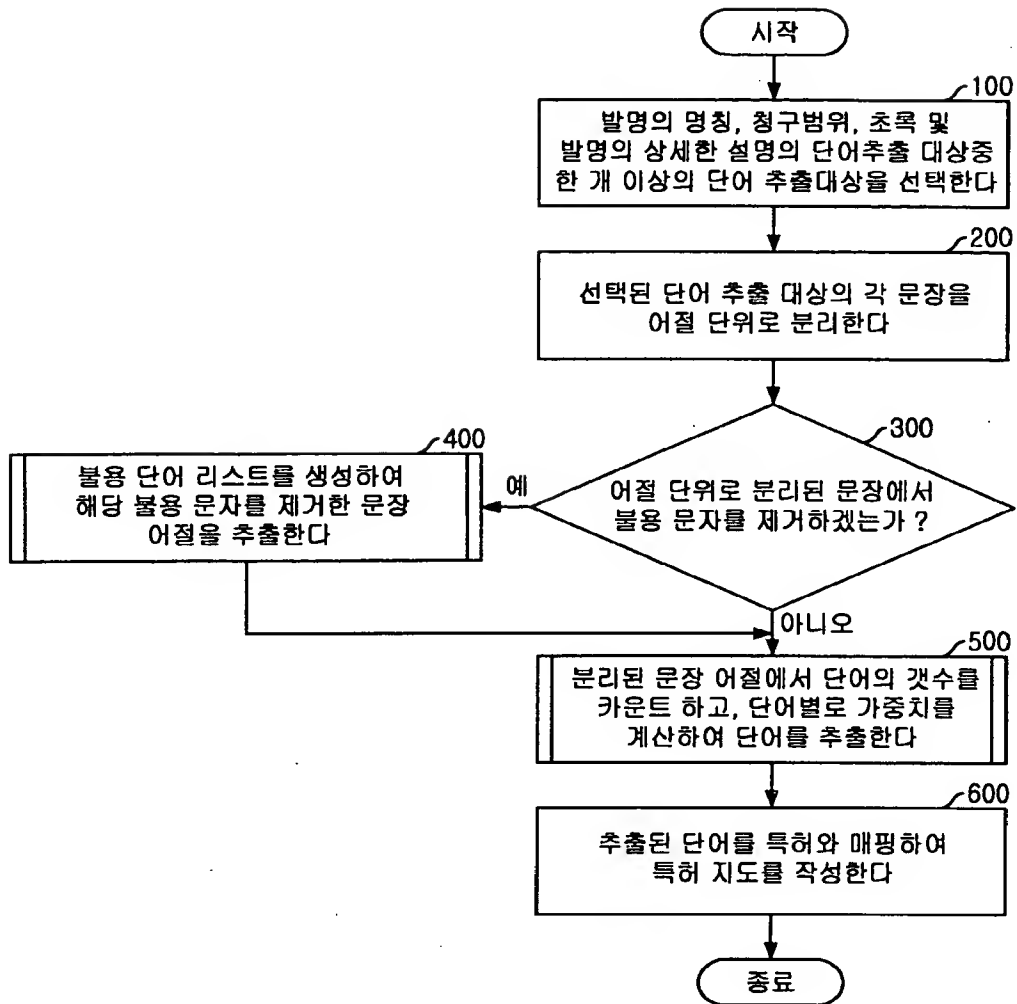
【도 2】



【도 3】

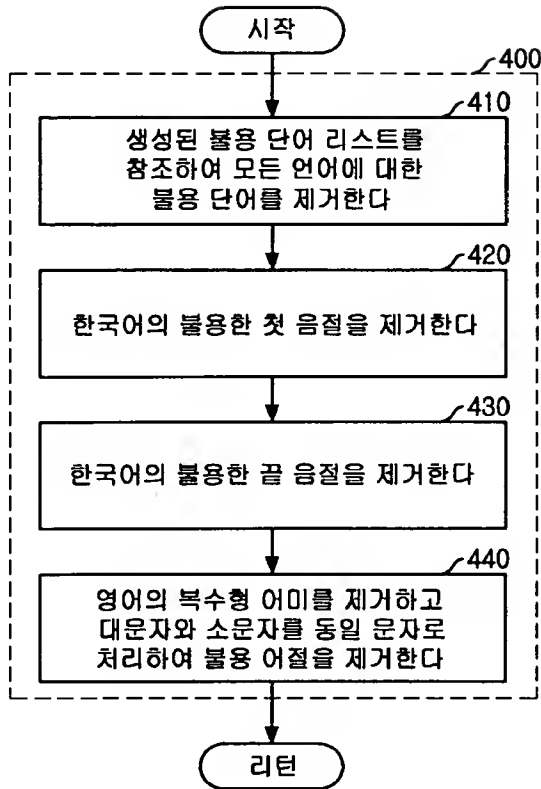


【도 4】





【도 5】



【도 6】

